CIBERSEGURIDAD

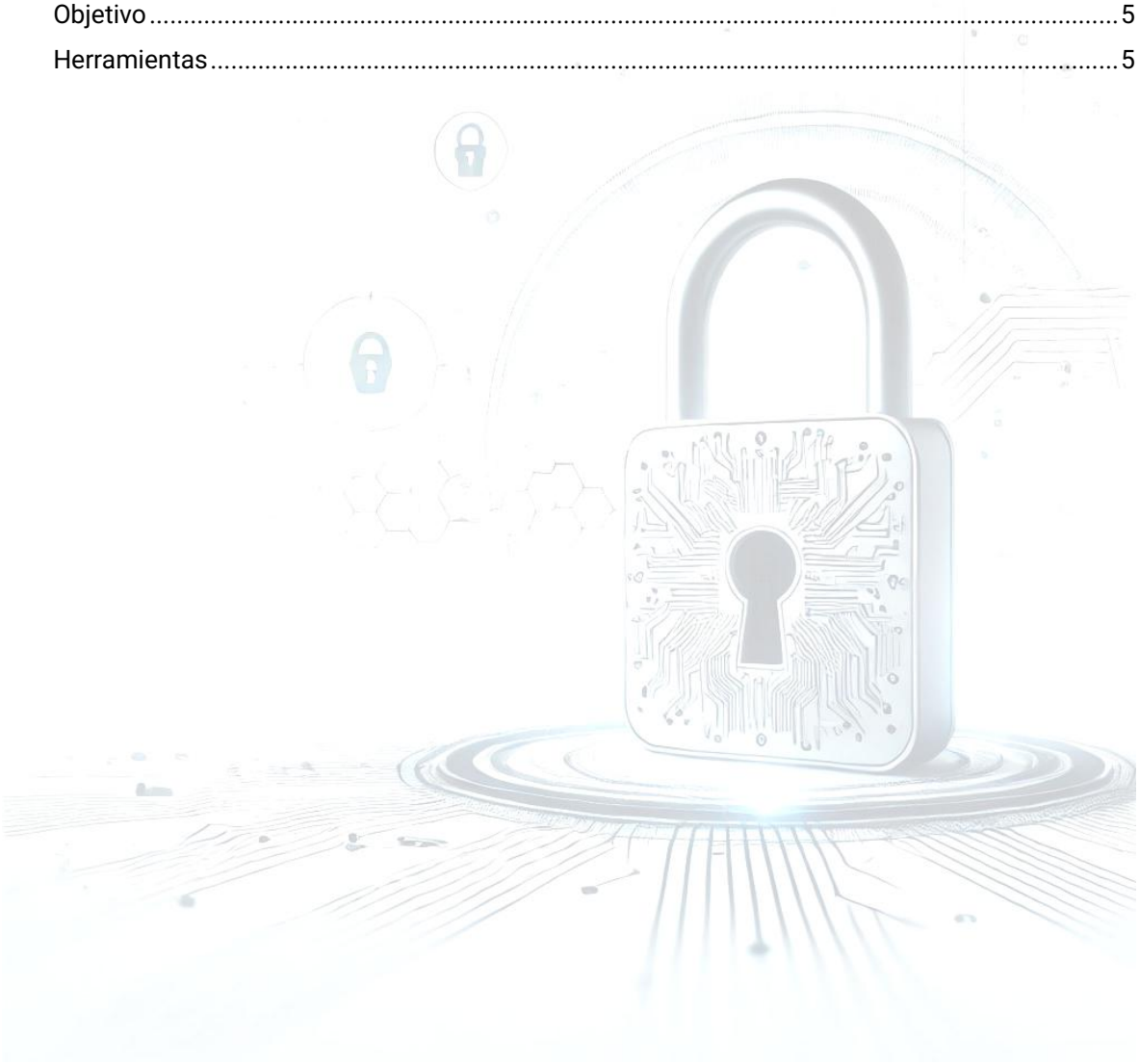
‘Bootcamp IX’

Informe Práctica Módulo Machine Learning y Ciberseguridad.

Maximiliano Dariel Altamirano.

Academia KeepCoding.

INDICE



INFORME

Descripción de la practica

Abordamos una problemática de clasificación para los datos del fichero dataset “cybersecurity_attacks.csv”.

Trabajaremos, en primera instancia, con la columna “Action Taken” como objetivo a predecir. Luego en el análisis extendido sumaremos las columnas “Severity Level” y “Attack Type”.

Gestionaremos los datos en el entorno JupyterLab.

Desarrollo de la práctica.

En el fichero “práctica_machine_learning.ipynb” realizamos el primer análisis y desarrollo cubriendo las expectativas básicas de la práctica. Segmentamos el análisis en:

1. Preparación de datos
2. Análisis exploratorio
3. Preprocesamiento/Generación de variables
4. División train/test
5. Modelado/Evaluación
6. Conclusiones

Entrenamos los modelos de ML:

- RandomForestClassifier
- LogisticRegression
- KNeighborsClassifier
- DecisionTreeClassifier

Este primer análisis se corresponde con los siguientes ficheros:

- práctica_machine_learning.ipynb
- cybersecurity_attacks.csv
- action_taken_test.csv
- action_taken_train.csv
- consigna_practica_final.pdf

PRACTICA EXTENDIDA

La entrega incluye un directorio “Extendido”, donde intentaremos evidenciar el análisis de la misma problemática con otro enfoque.

Realizamos una breve descripción de cada fichero:

- 01_análisis exploratorio.ipynb: intentaremos ser más minusiosos en la selección de columnas, exploración de datos y transformaciones. Del análisis realizado guardaremos la información en “02_clean_cybersecurity_attacks.csv” (datos ya transformados y seleccionados) y “03_codif_cybersecurity_attacks.csv” (datos codificados).

- 04_action_taken.ipynb: entrenaremos varios modelos de ML con la columna “Action Taken” como objetivo a predecir, descartando las columnas con menor importancia.

- 05_severity_level.ipynb: entrenaremos varios modelos de ML con la columna “Severity Level” como objetivo a predecir, descartando las columnas con menor importancia.

- 06_attack_type.ipynb: entrenaremos varios modelos de ML con la columna “Action Taken” como objetivo a predecir, descartando las columnas con menor importancia.

- 07_feature_importances.ipynb: considerando los datos obtenidos anteriormente, en este fichero planteamos trabajar con una cantidad de columnas reducidas, ahora seleccionando solo las columnas más importantes. Guardamos este análisis en el fichero “08_feature_importances.csv”.

En este punto decidimos utilizar solo dos modelos a entrenar, LogisticRegression y KNeighborsClassifier, ya que obtuvimos resultados equilibrados entre los conjuntos de datos “train” y “test”.

- 09_LogisticRegression.ipynb: entrenamos el modelo LogisticRegression y marcamos conclusiones con respecto a resultados anteriores

- 10_KNeighborsClassifier.ipynb: entrenamos el modelo KNeighborsClassifier y marcamos conclusiones.

RESUMEN

Objetivo

El informe pretende dejar en evidencia las herramientas y conocimientos adquiridos en el módulo entrenando un modelo de ML sobre un dataset puntual asignado con ese fin.

Herramientas

Utilizamos las siguientes herramientas:

- JupyterLab
- scikit-learn.org -web-
- motores de búsqueda -web-